Exploring the Effect of Gestures and Adaptive Tutoring on Children's Comprehension of L2 Vocabularies

Jan de Wit
TiCC*
Tilburg University
j.m.s.dewit@uvt.nl

Thorsten Schodde
Faculty of Technology, CITEC‖
Bielefeld University
tschodde@techfak.uni-bielefeld.de

Bram Willemsen
TiCC*
Tilburg University
b.willemsen@uvt.nl

Kirsten Bergmann
Faculty of Technology, CITEC‖
Bielefeld University
kirsten.bergmann@uni-bielefeld.de

Mirjam de Haas
TiCC*
Tilburg University
mirjam.dehaas@uvt.nl

Stefan Kopp
Faculty of Technology, CITEC‖
Bielefeld University
skopp@techfak.uni-bielefeld.de

Emiel Krahmer
TiCC*
Tilburg University
e.j.krahmer@uvt.nl

Paul Vogt
TiCC*
Tilburg University
p.a.vogt@uvt.nl

ABSTRACT

The L2TOR project explores the use of social robots for second language tutoring. This paper presents an experiment in preparation to investigate the effects of two educational scaffolding features (adaptation/personalization and iconic gestures), when used by a robot tutor, on children's comprehension of animal names in a foreign language. Participants will be children between the ages of four and five. The study is scheduled to take place in March 2017.

CCS CONCEPTS

•Computing methodologies → Cognitive robotics; Probabilistic reasoning; •Applied computing → Interactive learning environments; •Human-centered computing → *Empirical studies in HCI;*

KEYWORDS

Language tutoring; Assistive robotics; Education; Bayesian knowledge tracing; Human-robot interaction

ACM Reference format:
Jan de Wit, Thorsten Schodde, Bram Willemsen, Kirsten Bergmann, Mirjam de Haas, Stefan Kopp, Emiel Krahmer, and Paul Vogt. 2017. Exploring the Effect of Gestures and Adaptive Tutoring on Children's Comprehension of L2 Vocabularies. In *Proceedings of ACM HRI conference, Vienna, Austria, March 2017 (R4L workshop at HRI 2017)*, 6 pages.
DOI: 10.475/123_4

1 INTRODUCTION

The L2TOR project aims to design and develop a robot tutor capable of supporting children of four to five years old in the acquisition of a second language by interacting naturally with them in their social and referential environment through one-to-one tutoring interactions [1]. The robot used for the L2TOR project is the SoftBank Robotics NAO humanoid robot. The NAO robot is capable of speaking multiple languages, readily able to switch between them, which provides the possibility to vary the amount of the child's native language (L1) and the second language (L2) to be taught. Furthermore, the physical presence of a robot is shown to improve learning gains compared to its two-dimensional counterparts (e.g. Leyzberg et al. [12]).

This three-year project will result in an integrated lesson plan, which is expected to contain 24 lessons spanning three different domains (math, space, and mental state). To design these lessons, we analyze the way human tutors interact with children and investigate how different functionalities of the robot can be used to ensure a natural and productive interaction. In this paper, we propose an experiment to evaluate two such functionalities: personalized lessons by adjustment of the level of difficulty of the subject matter to the level of proficiency of the learner and the use of gestures when introducing the L2 words. We expect that both concepts will help to create and maintain common ground with the child, while also increasing comprehension and memorization potential of new words in the L2.

The importance of personalized adjustments in the robot's behavior has been substantiated in recent research showing that participants who received personalized lessons from a robot (based on heuristic skill assessment) outperformed others who received a non-personalized training [12]. Suboptimal robot behavior (e.g. distracting, incongruent or in other ways inappropriate social behavior) can even hamper learning [10].

One of the main advantages of choosing a humanoid robot as a tutor is its physical presence in the world, allowing for interactions similar to those between humans. Because of its anthropomorphic appearance, we tend to expect human-like communicative behavior

R4L workshop at HRI 2017, Vienna, Austria

DOI: 10.475/123_4

*Tilburg center for Cognition and Communication
‖Cluster of Excellence Cognitive Interaction Technology

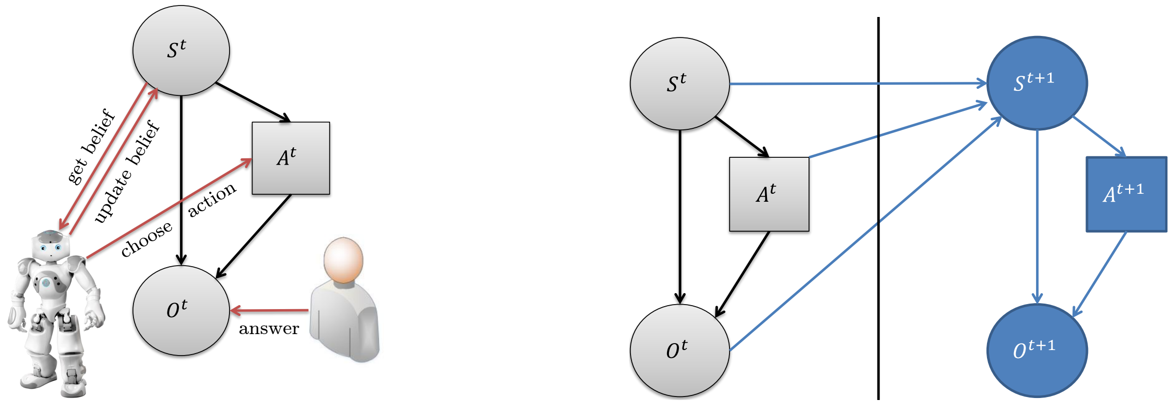

Figure 1: Dynamic Bayesian Network for BKT: With the current skill-belief the robot chooses the next skill S^t and action A^t for time step t (left). After observing an answer O^t from the learner, this observation together with action A^t and the previous skill-belief S^t are used to update the skill-belief S^{t+1} at time $t+1$ (right) [18].

from the robot, including proper use of non-verbal communication. Robots that perform gestures are perceived in a more positive way than those that use only speech [16].

In Section 2 we explain our previous work to evaluate adaptive learning, which is used as a starting point for the experiment described in this paper. We then introduce iconic gestures and describe how they could be used to increase learning gain in a human-robot tutoring context in Section 3, followed by our main research questions in Section 4. Section 5 outlines the design of the proposed experiment. We intend to start data collection in March 2017.

2 PREVIOUS WORK

2.1 Adaptive language tutoring with a robot

In previous work we developed a novel approach to personalize language tutoring in human-robot interaction [18]. This adaptive tutoring is enabled through a model of how tutors mentalize about learners – by keeping track of their knowledge state and by selecting the next tutoring actions based on their likely effects on the learner. This is realized via an extended model that combines knowledge tracing (of what the learner learned) with tutoring actions (of the tutor) in one causal probabilistic model. This allows for selecting skills and actions based on notions of optimality – here the desired learner's knowledge state as well as optimal task difficulty – to achieve this for a given skill.

The approach is based on Bayesian Knowledge Tracing (BKT) [4], a specific type of Dynamic Bayesian Networks (DBNs). The model consists of two types of variables, namely the *latent variables* representing the belief state of 'skills' to be acquired (e.g. whether a word has been learned or not) and the *observed variables* representing the observable information of the learning interaction (e.g. whether an answer was correct or not). In our proposed model, each latent variable can attain six discrete values, corresponding to six bins for the belief state (0%, 20%, 40%, 60%, 80%, 100%) representing whether a skill is mastered or not as a discretized probability distribution. That is, we reduce the complexity we would get through continuous latent variables but also attain more flexibility. The observed variables remain binary and still represent whether a learner's response is correct or not (see Figure 1). Moreover, the following update of the belief state of the skill, i.e. the skill-belief, at time $t+1$ is not only based on the previous skill-belief, but also on the chosen action and the previous observation at time t.

Based on this model, two types of decisions are made, (1) which skill would be the best to address next, and (2) the choice of action to address that skill. Regarding the former, we employ a heuristic maximizing the beliefs of all skills while balancing the single skill-beliefs among each other. This strategy is comparable to the vocabulary learning technique of *spaced repetition* as implemented, for instance, in the Leitner system [11]. Regarding the choice of action, the model enables the simulation of the impact each action has on a particular skill. To keep the model simple, the action space of the model consists of three different task difficulties (easy, medium, hard). Consider an example where the skill-belief appears relatively high, such that the skill is nearly mastered by the learner. In this case, a less challenging task would only result in a relatively minor benefit for the training of that skill. In contrast, if we assume the skill-belief to be rather low and a very difficult task is given, the student would barely be able to solve the task, likewise resulting in a smaller (or non-existent) learning gain. Instead, a task of adequate difficulty, not needlessly simple nor too complicated for the student to solve, will result in a higher learning gain [5]. This helps to position the robot as a capable instructor that uses these scaffolding techniques to help children acquire new skills beyond what they could have learned without help, by bringing them into the zone of proximal development (ZPD) [22].

2.2 Evaluation

When implemented in the robot language tutor, the model will enable the robot tutor to trace the learner's knowledge with respect to the words to be learned, to decide which skill (word) to teach next, and how to address the learning of this skill in a game-like tutoring interaction. For the experiment as described in [18], participants were asked to learn ten vocabulary items German – 'Vimmi'

(Vimmi is an artificial language that was developed to avoid associations with other known words or languages for language-related experiments [13]). The items included colors, shapes and the words 'big' and 'small'. During the game, the robot would introduce one of the Vimmi words. A tablet then displayed several images, one of which satisfied the Vimmi description (e.g. one object that is blue) and a number of distractors. The participant was then asked to select the image corresponding to the described item. Participants learned vocabulary items in one of two conditions, either in the condition with the adaptive model or in a non-adaptive (random) control condition. In the adaptive condition, the skill to be taught and the action to address the skill were chosen by the model as described above. Participants' performance was assessed with two measures: (1) learners' response behavior was tracked over the course of the training to investigate the progress of learning, and (2) a post-test was conducted on the taught vocabulary in the form of both L1-to-L2 translations and L2-to-L1 translations to assess participants' state of knowledge following the intervention.

Analysis of participants' response behavior over the course of training indicated that the participants learned the L2 words during the human-robot interaction (see [18] for more detailed results). Importantly, they learned more successfully with our adaptive model as compared to a randomized training. That is, the repeated trials addressing still unknown items as chosen by the adaptive model (until the belief state about these words equaled that of known items) outperformed the tutoring of the same material (same number of trials and items) but in randomized order. In the post-test, however, there was no significant difference across experimental conditions, despite a trend towards increased performance in the adaptive model conditions as compared to the controls.

3 ICONIC GESTURES

A growing body of evidence suggests that iconic gestures bear a great potential to enhance learners' memory performance for novel L2 words. Iconic gestures are movements that have a formal relation (in form or manner of execution) to the semantic content of the linguistic unit they describe [14]. In other words, the gesture elicits a mental image that relates strongly to the word or words that it links to. As an example, the word *bird* could be described by an iconic movement of stretching both arms sideways and moving them up and down, symbolizing the flapping of wings. The supporting effect of iconic gestures on L2 vocabulary learning by providing a congruent link between the words to be learned and gesture being observed or imitated has been shown in various studies (e.g. [6, 9, 13, 15, 19]). A recent overview of how gestures contribute to foreign language learning and possible explanations for this effect is given by Hald et al. [8]. Although they focus mainly on students *performing* or re-enacting the gestures, merely observing a gesture is shown to aid learning as well. Research conducted by Tellier [19] and De Nooijer et al. [6] investigated the role of gestures with respect to children and word learning. The effect of gestures is shown to depend on the students' gender, language background and existing experience with the L1 [15].

When considering the use of an artificial embodied agent as a tutor, the positive effects of gesturing seem to apply as well, as shown by Bergmann and Macedonia for a virtual tutor [2], and by Van Dijk et al. for a robotic tutor [20]. An additional benefit of implementing non-verbal behavior is to improve the way the robot is perceived, making it seem more human-like[17]. The challenge of mapping non-verbal behavior to the robot lies in the fact that each act needs to be carefully designed and choreographed to coincide with the corresponding word or sentence. There are limits to the degrees of freedom, the working space (i.e. the physical reach) and smoothness of motion that the robot has to offer. As an example, Figure 2 shows an attempt at making an iconic gesture for *rabbit*. The right arm has to take an unnatural position, which may result in an uncanny feeling for the observer. The NAO robot also has only three fingers and they cannot move independently, therefore finger-counting and similar subtle motions do not transfer to the robot without modification. The challenge lies in finding ways to work around these limitations, while still taking advantage of the added value of non-verbal communication. The gestures that were designed for this experiment have been exaggerated beyond what the human alternatives would look like. For example, when imitating a monkey the robot will bend its knees and shift its weight from side to side (see Figure 3).

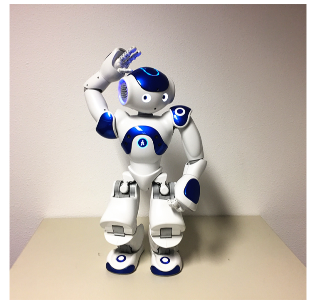

Figure 2: Attempt at showing an iconic gesture for a *rabbit*. The unnatural angle of the arm, positioning of the hand, and movement of the fingers, may lead to confusion and, consequently, adverse effects with respect to learning.

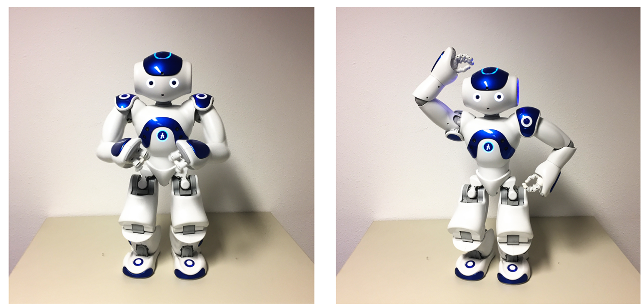

Figure 3: Stills of iconic gestures as depicted by the robot. Left: imitating a *chicken* by simulating the flapping of its wings; right: imitating a *monkey* by simulating the scratching of the head and armpit with the right and left extremities, respectively.

4 RESEARCH QUESTIONS

With the upcoming experiment we intend to answer two research questions. The first question relates to the previous work described in Section 2. We aim to investigate to what extent children will benefit from adaptive language tutoring. We hypothesize an increase in learning gain when children are taught words through an adaptive language tutoring system as compared to a non-adaptive (random) language tutoring system. We anticipate a difference in the exact words that are learned: in the adaptive condition, we expect children to learn those words that were the most challenging during training (having the most incorrect answers) because of the higher repetition rate of these words. In the random condition, the words learned might depend on other factors such as word complexity or attitude towards the animal described by the word.

Our second research question focuses on the effect of gestures on L2 comprehension for children. We hypothesize an increase in learning gain when target words are accompanied by (iconic) gestures during learning, as compared to the absence of gestures. Furthermore, we expect a reduced knowledge decay over time of the words in the gesture condition, similar to the discoveries by Cook et al. in the math problem solving domain with a human tutor [3]. We intend to investigate, using the retention test one week after the experiment, whether these findings carry over to the language learning domain with gestures performed by the robot. It should be noted that participants are not required but also not prohibited from using gestures during the experiment and pre- and post-tests. We are interested in seeing whether children will produce gestures spontaneously following training and, if so, to what extent these gestures will prove to be similar to the ones depicted by the robot.

5 PROPOSED EXPERIMENT

Following the two research questions, our experiment has a 2 (adaptive versus non-adaptive) x 2 (gestures versus no gestures) between-subjects design. We aim to recruit 80 participants, all native Dutch speaking children between the ages of four and five.

Although the proposed experiment is largely a replication of the experiment described in Section 2 and presented in [18], changes to the design had to be made to accommodate the younger participants, as the previous experiment was tailored to adults. Instead of the first interaction between the children and the robot taking place as part of the experiment, the robot will be introduced to the children in a group session the week prior to the experiment to build trust and rapport. We will refer to the robot by a proper name (Robin) and present a background story to stimulate a friendly and open attitude towards the robot [21].

Rather than teaching children the fictional Vimmi words, the target words are the English names of six animals: chicken, monkey, horse, spider, bird, and hippo (used instead of the more difficult hippopotamus). The number of words was reduced to six (from ten in the original experiment, see Schodde et al. [18]) to account for the lower word memory span of children [7], which should be around four words for children of age five. All target words have been selected based on the (varying degrees of) dissimilarity between the words in the L1 (Dutch) and the L2 (English) as well as the feasibility of designing suitable iconic gestures to be performed by the robot to depict the animals. We will conduct a pre-test

Figure 4: Mock-up of the training phase of the proposed experiment. Three animals appear on the tablet screen, one of which matches the animal picked by the robot. The robot asks the child in their L1 to point out the correct animal based on its name in the L2. In the gesture condition, as shown in this image, the robot performs the associated iconic gesture when mentioning the animal.

to verify that participants are familiar with all six target words in their L1 (Dutch) and to test the underlying assumption that participants have no prior knowledge of the target words in the L2 (English). This pre-test will be presented on a different computer screen than the one on which the game is played and without the robot being present, so that there is a clear distinction between this testing environment and the training (game) stage. On the computer screen, the participant will be presented with the pictures of all six animals, one by one. For each picture, the experimenter will ask the participant for the name of the animal in the L1. The computer will then show the pictures of all animals on the screen and name the animals, one after another, in the L2 in random order. Each time the child is prompted with a name in the L2, they are asked to pick the correct image for this animal from the six animals displayed.

The experimental setup uses a Microsoft Surface Pro 4 tablet and the SoftBank Robotics NAO robot. The robot plays a game of "I spy with my little eye", where it picks a certain animal displayed on the tablet screen and names it in the L2, after which the child is expected to tap the corresponding animal picture (see Figure 4). The experimenter inputs the name of the child, so that the robot can personally address the participant, and starts the game. After a brief explanation, the tablet will ask participants to indicate whether they understand the concept of the game. If they indicate that they do not, the experimenter will intervene to provide further explanations. The experiment can be stopped at any time via an experimenter-controlled control panel. Once the actual game commences, the experimenter pretends to be preoccupied so as to avoid participants actively looking for feedback.

In the adaptive learning condition the next target word to train is selected based on the knowledge model (i.e. skill-beliefs) of the participant. After each trial in which the robot exposes the child to one animal, this knowledge model is updated based on the responses of the child. The updated model is then used to select the next target word to be presented. In the random condition, target words are instead randomly presented. In total, there are thirty of these tasks, which means that in the random condition each target word is presented five times throughout the game. In the adaptive condition, the number of times each word occurs depends on how well the participant performs on that specific word, but all words are guaranteed to occur at least once. The previous experiment also consisted of a total of thirty tasks, but as there were ten target words there was less repetition. Reducing the number of words should avoid cognitive overload for the young participants while simultaneously offering more room for the adaptive system to learn the knowledge model of the child and repeat the words that require more attention.

A new addition to the experiment is a condition in which the robot will perform iconic gestures whenever one of the animal names is mentioned in English. These gestures were specifically designed for this experiment, where the robot tries to mimic the appearance or behavior of the animal. The timing of L2 word pronunciation is designed to occur close to the stroke of the gesture. This means that there is a pause in mid-sentence leading up to and after the L2 word, creating additional emphasis on the target. In the condition without gestures, a similar pause is introduced. The robot is set to "breathing mode" in all conditions, which means that it slowly moves its weight from one leg to the other while slightly moving its arms. This prevents the robot from being completely static while, in the gesture condition, reducing the surprise effect of an iconic gesture being triggered.

After thirty prompts to recognize the English animal names, the game finishes. The child is then presented with the post-test, again at the computer screen without the robot. The post-test is identical to the pre-test, except that we no longer test the animal names in the L1. The post-test is also identical across all conditions, so there are no gestures when the L2 words are presented. There are two different images for each animal, one of which will be used for the pre-test and post-test and the other for the game. The images of animals used in the pre-test and post-test feature the same character as those that appear during the game, but in a different pose. The pose in the set of images used during the game is designed to match the gesture that is shown by the robot, to avoid having a mismatch between both sources of visual information for some animal names, and a match for others [23]. For instance, for the word 'bird' the robot will display the act of flying by moving its arms up and down, therefore the bird in the image is also flying. The second set of images could feature the bird facing a different direction, sitting still. By using these two sets of images, we aim to test if children manage to map the English words not only to the specific corresponding image or mental representation of shape, but to the general concept of the animal. One week after the experiment we perform the post-test once again to measure the retention of the six target words.

To assess the iconicity of the gestures, we conducted a perception test with adult participants through an online survey. Participants ($N = 14$) were shown video recordings, one after another, of the six gestures performed by the robot. For each video, participants were asked to answer the question which animal the robot depicted by selecting the corresponding name of the animal in English from a list containing all six target words. The order in which the videos were shown, as well as the order of the items on the list containing the six animal names, was randomized for each participant. Results from the perception test are presented in Table 1. As can be seen from this confusion matrix, with an average accuracy of over 89 percent, participants were, on average, very accurate with respect to their predictions of the depicted gestures. In fact, for three of the six animals (monkey, horse, and bird), not a single mistake was made. With an average accuracy of just over 71 percent, the most ambiguous gestures were those representing the chicken and the hippo. However, it should be noted that participants typically came to realize they had made a mistake, after which they acted accordingly: for example, if a participant was shown the video recording of the chicken prior to that of the monkey and they had incorrectly selected *monkey* as their answer for the recording of the chicken, they would (now correctly) select *monkey* again as their answer when shown the recording of the monkey (we did not allow them to directly correct their past mistake). This implied correction, as well as the high accuracy on average, suggests that we may assume the gestures to be sufficiently iconic, especially as they will ultimately be presented in combination with the verbalization of the name of the associated animal.

In our analysis of the experimental results, we intend to measure performance (correct and incorrect answers) during the word training to monitor participants' progress over time in the different conditions. Time on task is measured both in the training "game" and in the post-test. In addition, we will make video recordings of the interaction with the robot for additional analyses (for instance to see if and at what rate children will mimic the robot's gestures). During the post-test we will record how many animals the children managed to correctly identify immediately after training. The retention test will measure decay of the newly attained words after one week.

Table 1: Confusion Matrix Perception Test

Actual \ *Perceived*	Chicken	Monkey	Horse	Spider	Bird	Hippo
Chicken	10	2	1	0	0	0
Monkey	0	14	0	0	0	0
Horse	0	0	14	0	0	0
Spider	0	0	1	13	0	0
Bird	0	0	0	0	14	0
Hippo	1	1	0	2	0	10

Note. Shaded cells indicate true positives.

6 CONCLUSION

The experiment proposed in this paper outlines two valuable topics of discussion for improving the interactions between children and robot, specifically in a tutoring setting. We aim to investigate how the order and frequency of presenting new words in the L2 for the purpose of second language learning can be personalized for each child to optimize learning gain, based on a probabilistic model that traces their knowledge of each word. Second, the experiment evaluates if the positive effect of performing iconic gestures for second language learning by human tutors carries over to the robot.

After running the experiment, future work includes incorporating our findings into the L2TOR project[1]. Adaptive learning will be integrated with the existing lesson plans, improving not only the way the content of each individual lesson is structured but also informing the choice of which words from previous lessons to repeat for better retention. If iconic gestures indeed prove to play a big part in learning and remembering new words, more of these non-verbal behaviors will be developed to accompany a greater number of (target) words and concepts. Furthermore, we will investigate the use of different types of gestures and explore ways of reducing the effort required to implement and orchestrate these gestures for robots. Our progress can be tracked via the project website[1].

7 ACKNOWLEDGMENTS

This work is partially funded by the H2020 L2TOR project (grant 688014), the Tilburg center for Cognition and Communication 'TiCC' at Tilburg University (Netherlands) and the Cluster of Excellence Cognitive Interaction Technology 'CITEC' (EXC 277), funded by the German Research Foundation (DFG), at Bielefeld University (Germany). We would like to thank all members of the L2TOR project for their valuable comments and suggestions that have contributed towards the design of the experiment.

REFERENCES

[1] Tony Belpaeme, James Kennedy, Paul Baxter, Paul Vogt, Emiel E.J. Krahmer, Stefan Kopp, Kirsten Bergmann, Paul Leseman, Aylin C. Küntay, Tilbe Göksun, Amit K. Pandey, Rodolphe Gelin, Petra Koudelkova, and Tommy Deblieck. 2015. L2TOR - Second Language Tutoring using Social Robots. In *Proceedings of the International Conference on Social Robotics (ICSR) 2015 WONDER Workshop.*

[2] Kirsten Bergmann and Manuela Macedonia. 2013. *A Virtual Agent as Vocabulary Trainer: Iconic Gestures Help to Improve Learners' Memory Performance.* Springer Berlin Heidelberg, Berlin, Heidelberg, 139–148. DOI: http://dx.doi.org/10.1007/978-3-642-40415-3_12

[3] Susan Wagner Cook, Zachary Mitchell, and Susan Goldin-Meadow. 2008. Gesturing makes learning last. *Cognition* 106, 2 (2008), 1047–1058.

[4] Albert T. Corbett and John R. Anderson. 1994. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction* 4, 4 (1994), 253–278. DOI: http://dx.doi.org/10.1007/BF01099821

[5] Scotty Craig, Arthur Graesser, Jeremiah Sullins, and Barry Gholson. 2004. Affect and learning: An exploratory look into the role of affect in learning with AutoTutor. *Journal of Educational Media* 29, 3 (2004), 241–250. DOI: http://dx.doi.org/10.1080/1358165042000283101

[6] Jacqueline A. de Nooijer, Tamara van Gog, Fred Paas, and Rolf A. Zwaan. 2013. Effects of imitating gestures during encoding or during retrieval of novel verbs on children's test performance. *Acta Psychologica* 144, 1 (2013), 173 – 179. DOI: http://dx.doi.org/10.1016/j.actpsy.2013.05.013

[7] Frank N. Dempster. 1981. Memory span: Sources of individual and developmental differences. *Psychological Bulletin* 89, 1 (1981), 63–100. DOI: http://dx.doi.org/10.1037/0033-2909.89.1.63

[8] Lea A. Hald, Jacqueline de Nooijer, Tamara van Gog, and Harold Bekkering. 2016. Optimizing Word Learning via Links to Perceptual and Motoric Experience. *Educational Psychology Review* 28, 3 (2016), 495–522. DOI: http://dx.doi.org/10.1007/s10648-015-9334-2

[9] Spencer D. Kelly, Tara McDevitt, and Megan Esch. 2009. Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and Cognitive Processes* 24, 2 (2009), 313–334. DOI: http://dx.doi.org/10.1080/01690960802365567

[10] James Kennedy, Paul Baxter, and Tony Belpaeme. 2015. The Robot Who Tried Too Hard: Social Behaviour of a Robot Tutor Can Negatively Affect Child Learning. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '15).* ACM, New York, NY, USA, 67–74. DOI: http://dx.doi.org/10.1145/2696454.2696457

[11] Sebastian Leitner. 1972. *So lernt man lernen. Der Weg zum Erfolg.* Herder, Freiburg.

[12] Dan Leyzberg, Samuel Spaulding, Mariya Toneva, and Brian Scassellati. 2012. The Physical Presence of a Robot Tutor Increases Cognitive Learning Gains. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society (CogSci 2012).* Curran Associates, Inc.

[13] Manuela Macedonia, Karsten Müller, and Angela D. Friederici. 2011. The impact of iconic gestures on foreign language word learning and its neural substrate. *Human Brain Mapping* 32, 6 (2011), 982–998. DOI: http://dx.doi.org/10.1002/hbm.21084

[14] David McNeill. 1985. So you think gestures are nonverbal? *Psychological Review* 92, 3 (1985), 350–371. DOI: http://dx.doi.org/10.1037/0033-295x.92.3.350

[15] Meredith L. Rowe, Rebecca D. Silverman, and Bridget E. Mullan. 2013. The role of pictures and gestures as nonverbal aids in preschoolers' word learning in a novel language. *Contemporary Educational Psychology* 38, 2 (2013), 109 – 117. DOI: http://dx.doi.org/10.1016/j.cedpsych.2012.12.001

[16] Maha Salem, Stefan Kopp, Ipke Wachsmuth, Katharina Rohlfing, and Frank Joublin. 2012. Generation and Evaluation of Communicative Robot Gesture. *International Journal of Social Robotics* 4, 2 (2012), 201–217. DOI: http://dx.doi.org/10.1007/s12369-011-0124-9

[17] Maha Salem, Katharina Rohlfing, Stefan Kopp, and Frank Joublin. 2011. A Friendly Gesture: Investigating the Effect of Multimodal Robot Behavior in Human-Robot Interaction. In *2011 RO-MAN.* Institute of Electrical and Electronics Engineers (IEEE). DOI: http://dx.doi.org/10.1109/roman.2011.6005285

[18] Thorsten Schodde, Kirsten Bergmann, and Stefan Kopp. 2017. Adaptive Robot Language Tutoring Based on Bayesian Knowledge Tracing and Predictive Decision-Making. In *Proceedings of HRI 2017.*

[19] Marion Tellier. 2008. The effect of gestures on second language memorisation by young children. *Gestures in Language Development* 8, 2 (2008), 219–235. DOI: http://dx.doi.org/10.1075/gest.8.2.06tel

[20] Elisabeth T. van Dijk, Elena Torta, and Raymond H. Cuijpers. 2013. Effects of Eye Contact and Iconic Gestures on Message Retention in Human-Robot Interaction. *International Journal of Social Robotics* 5, 4 (2013), 491–501. DOI: http://dx.doi.org/10.1007/s12369-013-0214-y

[21] Paul Vogt, Mirjam de Haas, Chiara de Jong, Peta Baxter, and Emiel Krahmer. in press. Child-Robot Interactions for Second Language Tutoring to Preschool Children. *Frontiers in Human Neuroscience* (in press).

[22] Lev Vygotsky. 1978. *Mind in society: The development of higher psychological processes.* Harvard University Press, Cambridge, MA.

[23] Rolf A. Zwaan, Robert A. Stanfield, and Richard H. Yaxley. 2002. Language Comprehenders Mentally Represent the Shapes of Objects. *Psychological Science* 13, 2 (2002), 168–171. DOI: http://dx.doi.org/10.1111/1467-9280.00430

[1] http://l2tor.eu